Early colorectal cancer treatment and outcomes in Scotland: real world evidence from national linked administrative data

Lemmon E[a,*], Hanna C[b], Diernberger K[a], Hall PS[a,c]

[a] *University of Edinburgh, Edinburgh Health Economics*
[b] *University of Glasgow, Cancer Research UK Clinical Trials Unit*
[c] *University of Edinburgh, Edinburgh Cancer Research Centre*

Abstract

Background: Colorectal cancer (CRC) is the third most common type of cancer in Scotland and the second leading cause of cancer death. Despite improvements in CRC survival over time, Scotland lags behind its UK and European counterparts. Linked administrative datasets can provide a real-world representation of current care as a basis for evaluation of new interventions or policies and to understand variation in care and outcomes. In this study, we aim to provide up to date, population level evidence on CRC treatment and survival in Scotland for patients treated with curative or palliative intent as a basis for the understanding of variation, and to provide data to underpin the evaluation of new treatments. **Methods**: We conducted a retrospective analysis of adults with an incident CRC registered on the Scottish Cancer Registry (ICD-10 codes C18-20) between January 2006 and December 2018. Data on patients with incident CRC was linked to hospital inpatient records allowing description of their demographic, diagnostic and treatment characteristics. For a curative cohort (n = 26,204) Cox-Proportional Hazards regression models were used to assess the factors affecting overall survival (OS) and CRC specific survival (CRCS). **Results**: Overall, 32,690 (73%) and 12,184 (27%) patients had a diagnosis of colon and rectal cancer respectively. Patients with rectal cancer had a higher comorbidity score compared to those with colon cancer (1.17 versus 1.04). Chemotherapy was used in 30% with rectal cancer and 42% with colon cancer. Radiotherapy use was 2% and 39% respectively. Five year OS (CRCS) within the curative cohort were 71% (81%) and 75% (82%) for patients with colon and rectal cancer respectively. After accounting for patient and tumour characteristics, the adjusted regression models show that several factors significantly affect all-cause and CRC-specific survival. **Conclusions**: National linked administrative datasets have the ability to provide real-world representation of the treatments and outcomes for patients with cancer. In a Scottish population of curative patients with CRC, there was significant variation in survival depending on sex and geography.

*Corresponding author. E-mail:elizabeth.lemmon@ed.ac.uk

1. INTRODUCTION

Colorectal cancer (CRC) is the third most common type of cancer for men and women in Scotland and the second leading cause of cancer death (Public Health Scotland, 2020; Information Services Division Scotland, 2019). It is projected that the number of new CRC cases in Scotland will increase by 43% by 2023-27 compared to 2008-12(Information Services Division Scotland, 2015).

Despite improvements in CRC survival over time, Scotland lags behind its UK and European counterparts (De Angelis et al., 2014; Ferlay et al., 2018). Previous research has shown that socio-economic deprivation and remoteness factors, for example distance from a cancer centre, are significantly associated with poorer survival outcomes in Scotland (Shack et al., 2007; Campbell et al., 2000) and in the UK (Lyratzopoulos et al., 2011; Smith et al., 2006; Møller et al., 2012). However, there are considerable differences in conclusions across studies and the mechanisms behind the observed relationships remain unclear (Paterson et al., 2014; Brewster et al., 2001; Hole and McArdle, 2002). Possible explanations for the differences in survival outcomes observed across geographies and patient cohorts include delayed presentation, stage at presentation, treatment and co-morbidity.

Research in this area, and ultimately patient outcomes, may be improved by utilising the vast amounts of administrative healthcare data that are collected routinely as part of the delivery of patient care (Lemmon et al., 2021). It provides an opportunity to generate evidence with a high degree of external validity, being entirely representative of current care (Connelly et al., 2016). Administrative records, unlike trial or observational research data, provide detailed evidence on whole populations, over extended periods of time. This enables research that is inclusive of patient groups who are traditionally harder to reach, while exploiting the longitudinal nature of the data offers insight into changes that occur over time.

As well as real-world representativeness, a greater breadth of information can be obtained by linkage between datasets. Previous research has shown that administrative data provide a sufficiently accurate source of information to describe patients with cancer and their outcomes (Goldsbury et al., 2012). At the same time, linkage between datasets can provide a richer characterisation of patient needs, which is essential to understand the mechanisms underlying differences in treatment and outcomes.

In England, linked administrative datasets have been used within the CRC context to investigate routes to diagnosis (Pearson et al., 2019); explore provider differences in post-colonoscopy recurrence rates (Burr et al., 2019); explain variation in treatment and outcomes (Morris et al., 2010; Taylor et al., 2021); describe management of disease (Birch et al., 2019) and much more.

In Scotland, there are very few CRC studies using linked administrative datasets to investigate CRC treatment and many have used administrative data from a single geographic area e.g. (Paterson et al., 2014; Robertson et al., 2004; Hole and McArdle, 2002). Further, published, population level statistics on survival tend to group stages and/or disease sites together, rather

than look at outcomes for patients who are treated on a curative pathway.

In this study, we contribute to existing research by providing up to date, population level evidence on CRC treatment and survival outcomes for patients who are diagnosed with CRC in Scotland. We use a newly established, unique CRC dataset, which links demographic data to the Scottish Cancer Registry and routine hospital admissions data. Full details of this dataset are described elsewhere (Hanna et al., 2021). In what follows, we firstly describe the demographic, diagnostic and treatment characteristics of patients with colon and rectal cancer. Secondly, for those patients treated with curative intent, we estimate their survival and assess the factors affecting their overall survival and CRC specific survival.

2. METHODS

The study population consisted of adults having an incident CRC registered on the Scottish Cancer Registry (ICD-10 codes C18, C19 and 20) between January 2006 and December 2018. Approval for the study was granted by the Public Benefit and Privacy Panel (PBPP) for health and social care, project number 1718-0026. The study meets the requirements for ethical approval set out by the East of Scotland NHS Research Ethics Service for the analysis of secondary National Services of Scotland (NSS) data (Ethics Number: Need to find this).

2.1. Data

We used the Scottish Cancer Registry (Scottish Morbidity Record 06 (SMR06)) to identify a cohort of patients diagnosed with CRC. National Records of Scotland (NRS) deaths data was used to provide survival outcomes and inpatient and day case hospital admissions data (SMR01) provided information on comorbidities and hospital use prior to cancer diagnosis. These three datasets were linked via a pseudonomysied patient identifier.

Scottish Cancer Registry (SMR06)

The Scottish Cancer Registry dataset includes information on all new diagnoses of cancer occurring within Scotland. This data is collected by Public Health Scotland (PHS) and contains diagnostic, staging and treatment information. Each SMR06 record for a patient corresponds to a unique cancer diagnosis for that individual.

In this study, we had access to all SMR06 records for patients who had a diagnosis of CRC (International Disease Classification 10th Revision (ICD-10) codes C18, C19 or C20) between January 2006 and December 2018.

Inpatient and day case admissions (SMR01)

The SMR01 dataset contains episode level data for all general/acute inpatient or day cases in Scottish NHS hospitals or Scottish NHS beds in non-NHS-institutions. This study used all patient SMR01 records between 1997 and 2018, for any patient present in the study SMR06 dataset.

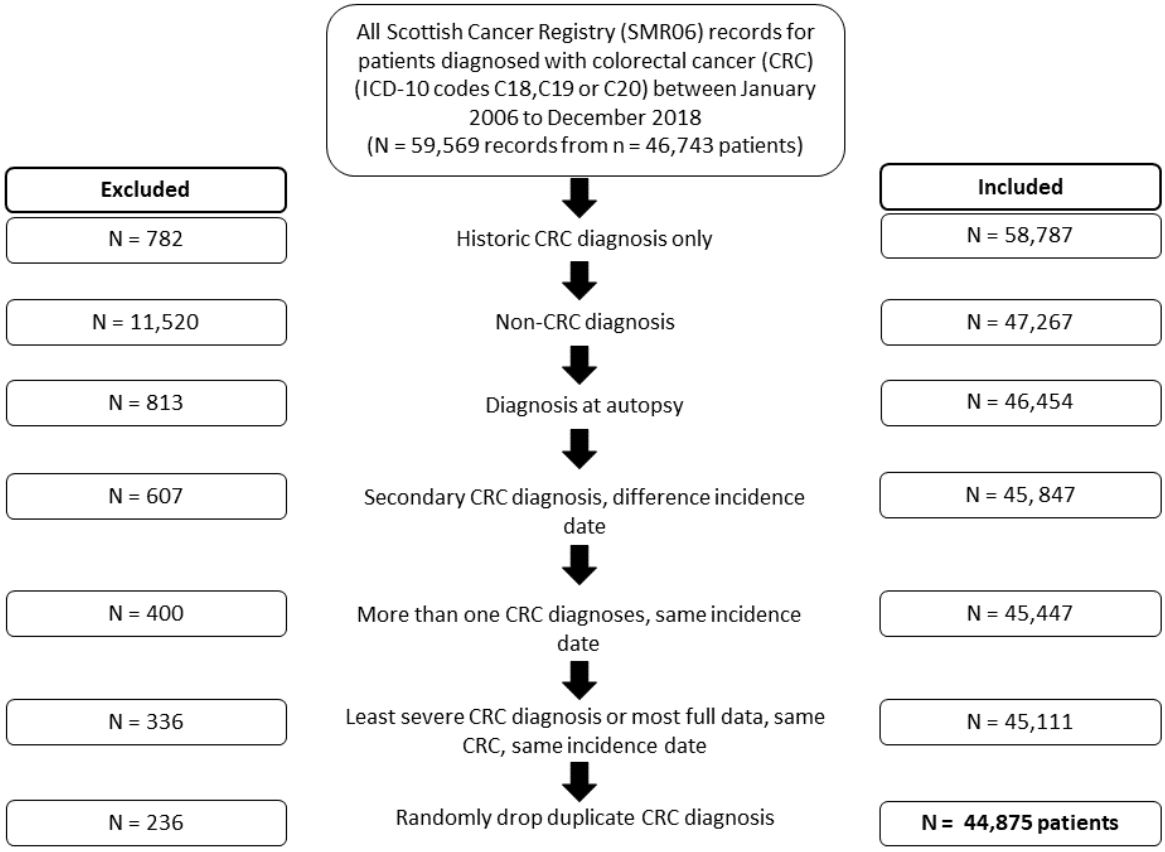

Figure 1: Cohort Derivation Flow Chart

National Records of Scotland (Deaths)

The NRS are responsible for the registration of all life events occurring in Scotland including births, deaths, marriages, civil partnerships and adoptions. For the purposes of this study, NRS vital events data on births and deaths was used to obtain patient date of birth, sex, date and cause of death.

2.2. Cohort derivation

The retrospective cohort was derived from the SMR06 database. Prior to matching patient records to their SMR01 and deaths records, a number of exclusion criteria were applied. These criteria are described in Fig. 1. In particular, patients were excluded if their diagnosis of CRC occurred before the study period using the flag for an historic CRC diagnosis (1.3% of records). Next, we excluded all non-CRC diagnoses (19.6% of records) and any CRC diagnosis that was diagnosed at autopsy (1.4% of records). Following this, using the incidence date, all secondary CRC diagnoses were excluded (1% of records). Further, patients were excluded in the event that they had more than one CRC diagnosis (i.e. different ICD-10 codes) with the same incidence date (less than 1% of records). In the event that a patient had more than one of the same CRC diagnosis (i.e. the same ICD-10 code) with the same incidence date, staging information

was used to exclude the least severe diagnosis and in the event that severity was identical, the records with the completion of most variables were kept (less than 1% of records). Finally, where incidence date, type of CRC, severity and data fullness were identical, duplicate records were randomly dropped.

Following exclusions, a total of N = 44,875 patients remained. This cohort was then linked to the SMR01 and deaths records to obtain pre-diagnosis co-morbidity, hospital admissions information and survival outcomes.

2.3. Descriptive analysis

The full cohort was characterised by descriptive statistics of their demographics, diagnosis and treatment. The variables included are described in Table 1 below.

Since the disease trajectories and treatment pathways are quite different for patients with rectal cancer (ICD.10 code C20) compared to patients with colon cancer (ICD-10 codes C18 and C19), analysis was carried out separately for these two disease sites.

2.4. Statistical analysis

All patients who were treated on a curative pathway were included in the survival analysis. We defined this curative cohort using a number of variables. Firstly, we removed those who did not undergo surgery or had an unknown surgery status (n = 11,506). Next, we removed those who had palliative or unknown therapy objectives (n = 4,618). Following this, we removed all patients who were diagnosed with stage IIII (n = 725) or unknown disease (n = 1,517). Finally, we removed patients where the chemotherapy and radiotherapy treatment variables were missing (n = 305). This resulted in a final curative cohort of 26,204 patients. Analysis of patients with metastatic or incurable CRC will be published separately.

We defined two survival end points: Overall survival (OS) and CRC survival (CRCS). OS was the interval between the date of diagnosis and the date of death, censored at the date of the most recent death observed in the data set in December 2018. CRCS was defined as the interval between the date of diagnosis and the date of death, where CRC was classified as the underlying cause of death. Those lost to follow up were censored in the same way as outlined for OS with the additional censoring of date of death for those who died from non-CRC related causes. Survival curves were estimated using the Kaplan-Meier method.

Finally, we conducted univariable and multivariable analysis to examine the factors associated with OS and CRCS using Cox-Proportional Hazards models. Explanatory variables were included in the multivariable model if they demonstrated statistical significance in the univariable models or where prior clinical justification was evident in the published literature.

Table 1: Demographics

Variable	Description
Demographics	
Sex	Binary sex indicator to indicate if the patient is male or female
Age	10 year age bands (18-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+) based on age at diagnosis
SIMD Quintile	Scottish Index of Multiple Deprivation Quintile from most deprived (1) to least deprived (5)
Urban/Rural	Binary indicator to indicate the rurality of the usual residence of the patient. Comes from the Scottish Government's six-fold urban/rural classification. Urban locations are defined as large urban areas, other urban areas or accessible small towns. Remote/rural are defined as remote small towns, accessible rural or remote rural.
Cancer Network	An indicator of the Managed Cancer Network (MCN)in which the patient was diagnosed. There are three MCNs in Scotland including South of Scotland Cancer Network (SCAN), West of Scotland Cancer Network (WoSCAN) and the North of Scotland Cancer Network (NoSCAN)
Diagnosis	
Stage (Duke's stage)	Indicates the extent of spread of the invasive tumour at diagnosis in terms of the pathological and/or clinical findings (Stage I, II, III, IIII or Unknown)
Method 1st detection	Categorical indicator to indicate how the tumour was first detected (Screening examination, incidental finding, clinical presentation, interval cancer or other, not known.
Year of diagnosis	Three year bands for the year of diagnosis (2006-08, 2009-11, 2012-14 and 2015-18)
Treatment	
Therapy objectives	Categorical indicator to indicate the treatment intent (curative, palliative or unknown)
Chemotherapy	Binary indicator to indicate if the patient has had systemic chemotherapy treatment
Radiotherapy	Binary indicator to indicate if the patient was treated with radiotherapy
Surgery	Binary indicator to indicate if the patient was treated with surgery
Inpatient episodes	Median number of hospital inpatient episodes or day cases in the year ore-diagnosis (derived from SMR01)
Quan morbidity score	Mean morbidity score (Quan et al., 2005) in the year pre-diagnosis (derived from SMR01) using Charlson indicators with Quan weights (excludes cancer and metastatic cancer).

3. RESULTS

3.1. Patient characteristics

The final study population included 44,875 patients. Patient characteristics are described in Table 2. Overall, 32,690 (73%) of patients had a diagnosis of colon cancer and 12,184 (27%) had a rectal cancer diagnosis. The majority of patients diagnosed with colon and rectal cancer were aged over 65.

The majority of patients lived in urban areas (around 76% for both cancers). WoSCAN is the largest MCN in Scotland and as expected, it accounts for the largest proportion of CRC

Table 2: Patient Characteristics

Total cohort N = 44,875					
		Colon		Rectal	
		N = 32,691		N = 12,184	
		n	**%**	n	%
Demographics					
Sex	**Male**	16,989	52.0	7,532	61.8
	Female	15,702	48.0	4,652	38.2
Age	**18-34**	265	0.8	91	0.7
	35-44	598	1.8	325	2.7
	45-54	2,360	7.2	1,227	10.1
	55-64	5,722	17.5	2,754	22.6
	65-74	9,923	30.4	3,776	31.0
	75-84	9,947	30.4	2,997	24.6
	85+	3,876	11.9	1,014	8.3
SIMD Quintile	**1**	6,296	19.3	2,368	19.4
	2	6,994	21.4	2,634	21.6
	3	7,002	21.4	2,516	20.7
	4	6,247	19.1	2,436	20.0
	5	6,152	18.8	2,230	18.3
Urban/Rural	**Urban**	25,018	76.5	9,242	75.9
	Remote or Rural	7,673	23.5	2,942	24.1
Cancer Network	**SCAN**	11,344	34.7	4,110	33.7
	WoSCAN	13,186	40.3	4,880	40.1
	NoSCAN	8,161	25.0	3,194	26.2
Mean Quan morbidity score	**5 years pre-diagnosis**	1.17	-	1.04	-
Diagnosis					
Stage	**I**	4,183	12.8	2,703	22.2
	II	8,911	27.3	2,271	18.6
	III	8,229	25.2	2,582	21.2
	IIII	6,929	21.2	1,876	15.4
	Unknown	4,439	13.6	2,752	22.6
Method 1st dectection	**Screening**	4,432	13.6	1,799	14.8
	Incidental finding	984	3.0	206	1.7
	Clinical presentation	26,911	82.3	10,065	82.6
	Interval cancer, other	152	0.5	32	0.3
	Unknown	212	0.6	82	0.7
Year of diagnosis	**2006-2008**	7,552	23.1	2,941	24.1
	2009-2011	8,298	25.4	3,094	25.4
	2012-2014	7,986	24.4	2,823	23.2
	2015-2018	8,855	27.1	3,326	27.3
Treatment					
Therapy objectives	**Curative**	19,311	59.1	7,585	62.3
	Palliative	11,404	34.9	3,659	30.0
	Unknown	1,976	6.0	940	7.7
Chemotherapy	**No**	22,723	69.5	6,922	56.8
	Yes/planned	9,654	29.5	5,136	42.2
	Not known	314	1.0	126	1.0
Radiotherapy	**No**	31,556	96.5	7,205	59.1
	Yes/planned	767	2.3	4,794	39.3
	Not known	368	1.1	185	1.5
Surgery	**No**	8,146	24.9	3,360	27.6
	Yes/planned	24,435	74.7	8,765	71.9
	Not known	110	0.3	59	0.5
Median no. inpatient episodes	**Within 1 year prior to diagnosis**	8	-	7	-
Mean Quan morbidity score	**5 years pre-diagnosis**	1.17	-	1.04	-

patients (40%), followed by SCAN (35% colon and 34% rectal) and NoSCAN (25% colon and 26% rectal).

In terms of the Scottish Index of Multiple Deprivation (SIMD), there is little variation across quintiles though for both cancer groups, a slightly lower proportion of patients come from the least deprived quintile (quintile five) compared to the most deprived quintile (quintile one).

Hospital admissions records five years prior to diagnosis show that patients diagnosed with colon cancer had a Quan co-morbidity score of 1.17, compared to 1.04 for patients diagnosed with rectal cancer (Quan et al., 2005).

3.2. Diagnosis

The majority of patients were diagnosed via clinical presentation, with approximately 14% of patients with colon cancer and 15% of patients with rectal cancer, diagnosed via screening. There were differences in stage of disease at diagnosis depending on disease site. For patients with rectal cancer, 21% had stage I disease and 19% had stage II disease at diagnosis compared to 13% and 27% respectively for patients with colon cancer. More patients with rectal cancer had an unknown stage of disease recorded at diagnosis (24% versus 15% for colon cancer). Diagnoses were approximately evenly distributed throughout the study period with between 23 and 25% of patients being diagnosed within each three year period and 27% in the final four years.

3.3. Treatment

The majority of patients were treated with curative intent. Specifically, 59% of patients diagnosed with colon cancer and 62% of patients with rectal cancer, with a smaller proportion treated with palliative intent (35% and 30% respectively). A small proportion of patients had unknown therapy objectives, 6% and 8% respectively. Accordingly, over 70% of patients with colon and rectal cancer underwent or had planned surgery.

The treatment for patients with colon cancer differs compared to those with rectal cancer. For the whole cohort, patients with a colon cancer diagnosis were less likely to receive chemotherapy (30%) compared to patients with rectal cancer (42%), and they were much less likely to receive radiotherapy (2% versus 39%).

On average, for both cancer types, patients had 11 inpatient or day case episodes in the year prior to their diagnosis.

3.4. Survival

Table 3 displays descriptive survival statistics for the curative cohort (n = 26,204). In total, 9,094 patients died between January 2006 and December 2018. Specifically, 36% (6,930) of patients with colon cancer died compared to 31% (2,164) of patients with rectal cancer. There were 5,040 CRC specific deaths. Around 19% of both patients with colon cancer (3,722) and rectal cancer (1,318) died from CRC specifically. Median survival was around 11 years for all

patients. For CRCS, median survival is not defined since more than 50% of patients had not died due to CRC at the end of the period.

Table 3: Survival Outcomes

Total cohort N = 26,204						
	All deaths			CRC deaths		
	All	**Colon**	**Rectal**	**All**	**Colon**	**Rectal**
Number of patients	26,204	19,284	6,920	26,204	19,284	6,920
Number of deaths	9,094	6,930	2,164	5,040	3,722	1,318
Total time at risk	142,813	103,590	39,223	142,813	103,590	39,223
Median survival	11.10	10.71	12.36	-	-	0-
3 year survival	82%	81%	85%	87%	87%	90%
5 year survival	72%	71%	75%	81%	81%	82%
10 year survival	54%	53%	57%	74%	74%	73%

Fig. 2 presents the Kaplan-Meier (KM) survival curves for OS and CRC. 5-year OS for patients diagnosed with colon cancer was 71% and 75% for patients with a rectal cancer diagnosis. This is compared to 81% and 82% respectively for CRCS. Fig. 3 and Fig. 4 plot the KM survival curves, again for OS and CRCS for both types of CRC, broken down by stage at diagnosis. Survival outcomes worsen in line with the stage of disease at diagnosis and the differences between the survival functions are statistically significant.

[INSERT FIGURE 2,3,4 HERE]

3.4.1. Cox-Proportional Hazard Regressions

Table 4 displays the results from the Cox-Proportional Hazard regressions for patients diagnosed with colon cancer. We show unadjusted (univariable) and adjusted (multivariable) hazard ratios (HR) and Confidence Intervals (CI) for OS and CRCS.

Within the time-frame of our analysis, the adjusted model for OS and CRCS show that females have a significantly lower risk of death compared to males (OS HR: 0.896, 95% CI: 0.850-0.945 and CRCS HR: 0.893, 95% CI: 0.831-0.961).

Table 4: Cox-Proportional Hazard Regressions: Patients with colon cancer, n =

		Overall Survival (OS)						**CRC Survival (CRCS)**					
		Unadjusted			**Adjusted**			**Unadjusted**			**Adjusted**		
Variable	**Category**	**HR**	**95 % CI**		**HR**	**95 % CI**		**HR**	**95 % CI**		**HR**	**95% CI**	
Sex	Male	1	-	-	1	-	-	1	-	-	1	-	-
	Female	**0.929**	**0.886**	**0.974**	**0.896**	**0.850**	**0.945**	0.949	0.889	1.012	**0.893**	**0.831**	**0.961**
Age	18-34	1	-	-	1	-	-	1	-	-	1	-	-
	35-44	1.119	0.655	1.911	0.770	0.414	1.435	1.187	0.695	2.026	0.720	0.372	1.396
	45-54	1.215	0.753	1.961	0.935	0.539	1.622	1.258	0.779	2.029	0.929	0.521	1.656
	55-64	1.540	0.965	2.458	1.173	0.686	2.006	1.350	0.846	2.155	1.049	0.596	1.847
	65-74	**2.219**	**1.394**	**3.530**	1.661	0.975	2.829	1.509	0.948	2.401	1.253	0.715	2.195
	75-84	**4.492**	**2.825**	**7.143**	**2.897**	**1.701**	**4.935**	**2.580**	**1.622**	**4.102**	**2.015**	**1.149**	**3.535**
	85+	**7.914**	**4.964**	**12.617**	**4.924**	**2.881**	**8.415**	**3.889**	**2.440**	**6.201**	**3.086**	**1.744**	**5.460**
Cancer Network	SCAN	1	-	-	1	-	-	1	-	-	1	-	-
	WOSCAN	**0.924**	**0.875**	**0.976**	0.947	0.888	1.010	**0.878**	**0.814**	**0.946**	**0.874**	**0.802**	**0.952**
	NOSCAN	0.987	0.929	1.049	1.021	0.953	1.094	0.989	0.912	1.073	1.005	0.915	1.104
SIMD	1	1	-	-	1	-	-	1	-	-	1	-	-
	2	**0.915**	**0.851**	**0.983**	0.924	0.851	1.003	0.958	0.868	1.057	0.966	0.865	1.079
	3	**0.838**	**0.779**	**0.901**	0.854	0.783	0.931	**0.887**	**0.803**	**0.979**	**0.870**	**0.773**	**0.979**
	4	**0.731**	**0.677**	**0.789**	**0.774**	**0.706**	**0.848**	**0.750**	**0.675**	**0.833**	**0.737**	**0.648**	**0.837**
	5	**0.704**	**0.652**	**0.761**	**0.750**	**0.687**	**0.820**	**0.765**	**0.689**	**0.850**	**0.749**	**0.663**	**0.846**
Stage	I	1	-	-	1	-	-	1	-	-	1	-	-
	II	**1.631**	**1.518**	**1.751**	**1.465**	**1.361**	**1.578**	**2.586**	**2.271**	**2.945**	**2.375**	**2.060**	**2.738**
	III	**2.409**	**2.241**	**2.590**	**2.711**	**2.488**	**2.954**	**5.765**	**5.080**	**6.542**	**5.991**	**5.161**	**6.955**
Urban/Rural	Urban	1	-	-	1	-	-	1	-	-	1	-	-
	Remote/rural	**0.930**	**0.879**	**0.984**	0.960	0.896	1.029	0.955	0.885	1.030	0.960	0.873	1.056
Year of diagnosis	2006-08	1	-	-	1	-	-	1	-	-	1	-	-
	2009-11	**0.843**	**0.795**	**0.894**	**0.932**	**0.871**	**0.996**	**0.874**	**0.805**	**0.948**	0.964	0.878	1.058
	2012-14	**0.736**	**0.688**	**0.787**	**0.845**	**0.782**	**0.912**	**0.746**	**0.682**	**0.816**	**0.822**	**0.742**	**0.911**
	2015-18	**0.697**	**0.640**	**0.759**	**0.808**	**0.734**	**0.889**	**0.711**	**0.639**	**0.792**	**0.771**	**0.684**	**0.870**
Inpatient episodes		**1.013**	**1.011**	**1.016**	**1.008**	**1.005**	**1.011**	**1.011**	**1.009**	**1.014**	**1.011**	**1.008**	**1.014**
Quan morbidity score		**1.212**	**0.687**	**1.459**	**1.077**	**1.061**	**1.093**	**1.053**	**1.035**	**1.072**	**0.944**	**0.922**	**0.967**
Chemotherapy	No	1	-	-	1	-	-	1	-	-	1	-	-
	Yes/Planned	**0.721**	**0.682**	**0.763**	**0.708**	**0.653**	**0.769**	**1.220**	**1.140**	**1.305**	**0.787**	**0.712**	**0.871**
Radiotherapy	No	1	-	-	1	-	-	1	-	-	1	-	-
	Yes/Planned	**1.829**	**1.513**	**2.212**	**2.508**	**2.018**	**3.115**	**2.685**	**2.196**	**3.282**	**2.847**	**2.258**	**3.591**

Significance at the 5% level denoted by bold text

Table 5: Cox-Proportional Hazard Regressions: Patients with rectal cancer, n =

		Overall Survival (OS)						CRC Survival (CRCS)					
		Unadjusted			Adjusted			Unadjusted			Adjusted		
Variable	**Category**	**HR**	**95 % CI**		**HR**	**95 % CI**		**HR**	**95 % CI**		**HR**	**95% CI**	
Sex	Male	1	-	-	1	-	-	1	-	-	1	-	-
	Female	**0.807**	**0.739**	**0.883**	**0.780**	**0.703**	**0.865**	**0.792**	**0.706**	**0.889**	**0.756**	**0.659**	**0.867**
Age	18-34	1	-	-	1	-	-	1	-	-	1	-	-
	35-44	1.525	0.713	3.263	1.566	0.612	4.003	1.229	0.580	2.605	1.342	0.539	3.343
	45-54	1.024	0.496	2.114	1.089	0.442	2.682	0.784	0.385	1.597	0.862	0.359	2.066
	55-64	1.279	0.627	2.607	1.137	0.467	2.767	0.932	0.465	1.871	0.888	0.376	2.100
	65-74	1.762	0.867	3.583	1.501	0.619	3.637	0.961	0.480	1.924	0.947	0.402	2.233
	75-84	**3.289**	**1.618**	**6.686**	**2.596**	**1.071**	**6.291**	1.691	0.844	3.389	1.705	0.723	4.022
	85+	**6.520**	**3.170**	**13.410**	**5.356**	**2.185**	**13.128**	**3.375**	**1.646**	**6.919**	3.923	1.617	9.519
Cancer Network	SCAN	1	-	-	1	-	-	1	-	-	1	-	-
	WOSCAN	0.958	0.867	1.059	1.028	0.913	1.156	0.935	0.823	1.064	1.030	0.879	1.206
	NOSCAN	1.073	0.966	1.192	**1.140**	**1.005**	**1.293**	1.117	0.977	1.277	**1.186**	**1.006**	**1.398**
SIMD	1	1	-	-	1	-	-	1	-	-	1	-	-
	2	**0.875**	**0.768**	**0.997**	0.897	0.770	1.044	0.882	0.743	1.046	0.878	0.714	1.081
	3	0.888	0.780	1.012	1.047	0.896	1.223	0.919	0.776	1.089	1.054	0.857	1.297
	4	**0.794**	**0.696**	**0.906**	0.889	0.754	1.049	0.854	0.720	1.012	0.872	0.702	1.084
	5	**0.743**	**0.647**	**0.853**	**0.844**	**0.714**	**0.997**	**0.794**	**0.664**	**0.948**	0.855	0.687	1.065
Stage	I	1	-	-	1	-	-	1	-	-	1	-	-
	II	**1.542**	**1.385**	**1.718**	**1.368**	**1.212**	**1.544**	**2.099**	**1.799**	**2.449**	**1.732**	**1.446**	**2.074**
	III	**2.060**	**1.854**	**2.290**	**2.364**	**2.075**	**2.693**	**3.440**	**2.975**	**3.977**	**3.509**	**2.922**	**4.214**
Urban/Rural	Urban	1	-	-	1	-	-	1	-	-	1	-	-
	Remote/rural	0.984	0.893	1.084	0.915	0.809	1.035	1.026	0.907	1.161	0.969	0.824	1.140
Year of diagnosis	2006-08	1	-	-	1	-	-	1	-	-	1	-	-
	2009-11	0.950	0.856	1.054	1.034	0.914	1.169	0.889	0.777	1.017			
	2012-14	**0.800**	**0.708**	**0.905**	0.916	0.792	1.061	**0.785**	**0.674**	**0.913**	0.864	0.718	1.039
	2015-18	**0.674**	**0.570**	**0.798**	**0.824**	**0.677**	**1.001**	**0.732**	**0.601**	**0.891**	0.837	0.664	1.055
Inpatient episodes		**1.010**	**1.008**	**1.013**	**1.006**	**1.003**	**1.009**	**1.010**	**1.008**	**1.013**	**1.007**	**1.002**	**1.012**
Quan morbidity score		**1.218**	**1.193**	**1.243**	**1.124**	**1.094**	**1.156**	**1.106**	**1.073**	**1.139**	1.035	0.993	1.079
Chemotherapy	No	1	-	-	1	-	-	1	-	-	1	-	-
	Yes/Planned	**0.885**	**0.812**	**0.965**	**0.743**	**0.651**	**0.848**	**1.285**	**1.153**	**1.432**	**0.831**	**0.703**	**0.981**
Radiotherapy	No	1	-	-	1	-	-	1	-	-	1	-	-
	Yes/Planned	**1.228**	**1.126**	**1.340**	**1.506**	**1.345**	**1.685**	**1.639**	**1.470**	**1.826**	**1.843**	**1.601**	**2.122**

Significance at the 5% level denoted by bold text

Further, compared to those aged between 18 and 35, those aged 75 and above have a significantly higher likelihood of dying from CRC and this risk increases with age. The adjusted CRCS HR for those aged 75-84 is 2.015 (95% CI: 1.149-3.535) and for those aged 85+, the CRCS HR is 3.086 (95% CI: 1.744-5.460).

In terms of geography, cancer network has almost no effect on surival with the exception of WoSCAN where patients have significantly reduced risk of death from CRC relative to patients in SCAN (HR: 0.874, 95% CI: 0.802-0.952). The adjusted models show that living in a remote or rural area has no influence on risk of death.

Furthermore, the unadjusted OS results show that the risk of dying decreases with lower levels of deprivation. After adjusting for all other factors, no differences were observed between the 2nd and the 3rd quintiles but the effects remain significant for the two least deprived quintiles. Specifically, the HR for those living in the two least deprived areas (quintiles four and five) are 0.774 (95% CI: 0.706-0.848) and 0.75 (95% CI: 0.687-0.820) respectively. A similar trend was observed for CRCS, with the adjusted model showing a lower risk of death with reduced deprivation.

In the adjusted models, compared to those with stage I disease, the HR for risk of death from any cause for those with sage III is 2.711 (95% CI: 2.488-2.954) and for CRC specific deaths, the HR is 5.991 (95% CI: 5.161-6.955).

The year of diagnosis also has a significant influence on the risk of dying. The adjusted HRs show a significant reduction in risk of death overtime and thus an improvement in survival. For example, the risk of death from CRC was lower if a patient was diagnosed in 2015-18 compared to if they were diagnosed in 2006-08 (HR: 0.771, 95% CI: 0.684-0.870).

In terms of inpatient episodes within the year pre-diagnosis, all models show a significant increase in the likelihood of death as the number of hospital episodes increases. In the adjusted model for OS, the HR is 1.008 (95% CI: 1.005-1.011) and in the adjusted model for CRCS, the HR is 1.011 (95% CI: 1.008-1.014).

As for comorbidity, the models for OS show an increased risk of death as co-morbidity increases, even after accounting for other factors. Specifically, the HR on the Quan co-morbidity score variable is 1.077 (95% CI: 1.061-1.093). In contrast, in the models for CRCS, the risk of death falls as co-morbidity increases. That is, the HR is less than 1 (HR: 0.944, 95% CI: 0.922-0.967).

The chemotherapy variable indicates that receipt of chemotherapy is associated with a significantly reduced risk of death and this effect is larger for OS. In the adjusted model for OS, the HR for receipt of chemotherapy is 0.708 (95% CI: 0.653-0.769) compared to 0.787 (95% CI: 0.712-0.871) for CRCS.

The radiotherapy indicator is greater than one and significant across all of the models suggesting that receipt of radiotherapy is associated with a significantly increased risk of death.

Table 5 displays the Cox-Proportional Hazard results for patients diagnosed with rectal can-

cer. The results from the adjusted models for OS and CRCS show that females have a lower risk of death compared to males (OS HR: 0.780, 95% CI: 0.703-0.865, CRCS HR: 0.756, 95% CI: 0.659-0.867. After controlling for other factors, age has no impact on CRCS for patients with rectal cancer.

The adjusted models for patients diagnosed with rectal cancer suggest that compared to patients diagnosed in SCAN, there is no significant difference in risk of death compared to patients diagnosed in WOSCAN. However, rectal cancer patients diagnosed in NOSCAN have a higher risk of death. The HRs are 1.14 (95% CI: 1.005-1.293) and 1.186 (95% CI: 1.006-1.398) for OS and CRCS respectively.

Upon adjusting for other factors, the OS model suggests that the risk of death from any cause is significantly lower for the least deprived area (HR: 0.844, 95% CI: 0.714-0.997) but no significant differences between deprivation quintiles are observed for CRCS.

Stage of disease at diagnosis is associated with an increasing risk of death. The HRs in the adjusted models are 2.364 (95% CI: 2.075-2.693) and 3.509 (95% CI: 2.922-4.214) for OS and CRCS respectively.

Survival outcomes have improved over time, however, these effects are mostly insignificant.

An increase in the number of hospital inpatient episodes during the year pre-diagnosis is associated with a significant increase in the risk of death from any cause (adjusted HR: 1.124, 95% CI: 1.094-1.156), however we find no significant effect in the model for CRCS.

In addition, rectal cancer patients who received chemotherapy had a reduced risk of death (HR: 0.743, 95% CI: 0.651-0.848) and a reduced risk of death from CRC specifically (HR: 0.831, 95% CI: 0.703-0.981).

Finally, the radiotherapy indicator displays HRs that are greater than unity and significant across all model specifications, suggesting that radiotherapy is associated with an increased risk of death.

4. DISCUSSION

This paper used a national linked administrative dataset to provide up to date, real world evidence on the treatment and survival outcomes for patients diagnosed with CRC. We described the demographic, diagnostic and treatment characteristics of patients with colon cancer and rectal cancer. Further, for those patients treated on a curative pathway, we assessed the factors affecting patients overall and CRC specific survival.

Cox-proportional hazard models for a curative cohort of patients with colon or rectal cancer confirm that, older individuals are at increased risk of death from CRC specifically. Further, the stage at diagnosis influences survival, with those diagnosed at later stages having poorer survival compared to those diagnosed at the earliest stage. The adjusted models for OS show that patients diagnosed with stage III colon cancer have a threefold risk of death when compared to those diagnosed at stage I. Similarly, patients diagnosed with rectal cancer at stage III have a

risk of death that is two times higher compared to those at stage I.

The models show that survival has improved consistently since 2006-08. Furthermore, the models also find that a higher number of inpatient episodes in the year pre-diagnosis is associated with significantly poorer survival outcomes. The models also include a control for the level of co-morbidity of the patient. This variable shows that a higher co-morbidity score is associated with better CRC survival outcomes in patients diagnosed with colon cancer but has no impact on CRC survival for patients with rectal cancer.

In terms of treatment, the models find that patients who received chemotherapy tended to do significantly better compared to those who did not. However, for radiotherapy the models suggest that those who received radiotherapy had worse survival outcomes, both in terms of OS and CRCS. Whilst this appears somewhat counter intuitive, we believe that the radiotherapy variable may be picking up variation in tumour characteristics that have not been included in the models. In particular, for patients with colon cancer, radiotherapy would not be administered as part of standard care. It may be given to patients with colon cancer in circumstances where a patient is not well enough to receive other forms of treatment for the purposes of disease palliation and usually at the point of disease recurrence, rather than for treatment of primary disease. Furthermore, patients with early stage rectal cancer are often treated with surgery alone, whereas patients with large primary tumours, a high nodal burden or other evidence of locally advanced disease (e.g. threatened CRM) on pre-operative imaging will often receive radiation treatment in the neo-adjuvant setting as part of multi-modality treatment approach. Post-operative radiotherapy may be given if poor prognostic disease indicators are found on pathological examination and if radiation was not given pre-operatively. Therefore, notwithstanding the recent interest in organ preservation strategies, in general, for patients with rectal cancer treated with radical intent, those receiving radiotherapy will often have more aggressive, locally advanced disease compared to those treated with surgery alone. Overall, the radiotherapy indicator may be picking up a group of comorbid patients with high risk disease. Additionally, disease recurrence is not accounted for in our dataset and therefore we were not able to account for this possibility. Nevertheless, this finding requires further investigation.

However, the models also show a worrying result in that they suggest significant regional variation in outcomes between the three MCNs. In particular, patients diagnosed with colon cancer in WoSCAN appear to have better CRC survival outcomes compared to patients in SCAN. At the same time, patients diagnosed with rectal cancer in NoSCAN appear to have significantly poorer survival outcomes, both all cause and CRC specifically, compared to those in SCAN. These differences suggest potentially inequitable outcomes for patients in an NHS that is designed to ensure equal access to quality care. This finding requires further investigation to understand why we see significant variation in survival between the MCNs, despite controlling for needs and treatment variables. The models also account for location via the urban/rural indicator. In contrast to previous literature, the models do not find any evidence that

rurality has a significant impact on survival (Campbell et al., 2000).

Another concerning finding from our study is that among patients with colon cancer, the level of deprivation in the local area is significantly associated with poorer survival outcomes, even after accounting for clinical factors and patient need. This finding is consistent with some previous research carried out in Scotland (Shack et al., 2007). The reasons for poorer outcomes in the most deprived areas in Scotland is potentially due to later presentation or delay in treatment, however recent evidence from one MCN in Scotland finds that there is no association between deprivation and these factors (Paterson et al., 2014). Another possible explanation for the observed differences could be other co-morbidities present within more deprived populations, though we attempt to control for this effect via measures from hospital admissions data. Once again, further investigation into these differences is warranted. In particular, an investigation into type, length and timing of treatment.

In addition, the models find evidence that women have a significant survival advantage compared to men. This result is consistent with previous evidence from Scotland and around the world (McArdle et al., 2003; Yang et al., 2017). The explanation for these differences remains unclear but might be explained partly by endogenous factors such as genetics or hormones, particularly those present in younger females (Majek et al., 2013).

In summary, we have used a nationally linked administrative data set to retrospectively explore patterns in treatment and outcomes for CRC patients in Scotland. We have demonstrated that Scotland's unique data linkage infrastructure can accommodate linkage between demographic records, cancer registry and hospital admissions data, providing a fuller picture of the needs of CRC patients. We have identified a number of areas which require further research including regional and socio-demographic differences in outcomes.

Key next steps in our analysis will be to investigate these areas further and utilise a number of additional administrative data sets that have been linked to the registry data for the purposes of this project (Hanna et al., 2021). These include detailed information on chemotherapy prescribing and cancer audit data.

References

Birch, R.J., Taylor, J.C., Downing, A., Spencer, K., Finan, P.J., Audisio, R.A., Carrigan, C.M., Selby, P.J., Morris, E.J., 2019. Rectal cancer in old age–is it appropriately managed? evidence from population-based analysis of routine data across the english national health service. European Journal of Surgical Oncology 45, 1196–1204.

Brewster, D.H., Thomson, C.S., Hole, D.J., Black, R.J., Stroner, P.L., Gillis, C.R., 2001. Relation between socioeconomic status and tumour stage in patients with breast, colorectal, ovarian, and lung cancer: results from four national, population based studies. Bmj 322, 830–831.

Burr, N.E., Derbyshire, E., Taylor, J., Whalley, S., Subramanian, V., Finan, P.J., Rutter, M.D., Valori, R., Morris, E.J., 2019. Variation in post-colonoscopy colorectal cancer across colonoscopy providers in english national health service: population based cohort study. bmj 367.

Campbell, N., Elliott, A., Sharp, L., Ritchie, L., Cassidy, J., Little, J., 2000. Rural factors and survival from cancer: analysis of scottish cancer registrations. British Journal of Cancer 82, 1863–1866.

Connelly, R., Playford, C.J., Gayle, V., Dibben, C., 2016. The role of administrative data in the big data revolution in social science research. Social Science Research 59, 1–12.

De Angelis, R., Sant, M., Coleman, M.P., Francisci, S., Baili, P., Pierannunzio, D., Trama, A., Visser, O., Brenner, H., Ardanaz, E., et al., 2014. Cancer survival in europe 1999–2007 by country and age: results of eurocare-5—a population-based study. The lancet oncology 15, 23–34.

Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., Bray, F., 2018. Cancer incidence and mortality patterns in europe: estimates for 40 countries and 25 major cancers in 2018. European journal of cancer .

Goldsbury, D.E., Armstrong, K., Simonella, L., Armstrong, B.K., O'Connell, D.L., 2012. Using administrative health data to describe colorectal and lung cancer care in new south wales, australia: a validation study. BMC health services research 12, 1–9.

Hanna, C.R., Lemmon, E., Ennis, H., Jones, R.J., Hay, J., Halliday, R., Clark, S., Morris, E., Hall, P., 2021. Creation of the first national linked colorectal cancer dataset in scotland: prospects for future research and a reflection on lessons learned. International Journal of Population Data Science 6.

Hole, D., McArdle, C., 2002. Impact of socioeconomic deprivation on outcome after surgery for colorectal cancer. British journal of surgery 89, 586–590.

Information Services Division Scotland, 2015. Cancer incidence projections for scotland. https://beta.isdscotland.org/find-publications-and-data/conditions-and-diseases/cancer/cancer-incidence-projections-for-scotland/.

Information Services Division Scotland, 2019. Cancer mortality in scotland (2018). https://www.isdscotland.org/Health-Topics/Cancer/Publications/2019-10-29/2019-10-29-Cancer-Mortality-Summary.pdf.

Lemmon, E., Hanna, C.R., Hall, P., Morris, E., 2021. Health economic studies of colorectal cancer and the contribution of administrative data: A systematic review. European Journal of Cancer Care -.

Lyratzopoulos, G., Barbiere, J., Rachet, B., Baum, M., Thompson, M., Coleman, M., 2011. Changes over time in socioeconomic inequalities in breast and rectal cancer survival in england and wales during a 32-year period (1973–2004): the potential role of health care. Annals of Oncology 22, 1661 – 1666. URL: http://www.sciencedirect.com/science/article/pii/S0923753419384443, doi:https://doi.org/10.1093/annonc/mdq647.

Majek, O., Gondos, A., Jansen, L., Emrich, K., Holleczek, B., Katalinic, A., Nennecke, A., Eberle, A., Brenner, H., Group, G.C.S.W., et al., 2013. Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in germany. PloS one 8, e68077.

McArdle, C., McMillan, D., Hole, D., 2003. Male gender adversely affects survival following surgery for colorectal cancer. British Journal of Surgery 90, 711–715.

Møller, H., Sandin, F., Robinson, D., Bray, F., Klint, Linklater, K.M., Lambert, P.C., Påhlman, L., Holmberg, L., Morris, E., 2012. Colorectal cancer survival in socioeconomic groups in england: variation is mainly in the short term after diagnosis. European Journal of Cancer 48, 46–53.

Morris, E., Forman, D., Thomas, J., Quirke, P., Taylor, E., Fairley, L., Cottier, B., Poston, G., 2010. Surgical management and outcomes of colorectal cancer liver metastases. British Journal of Surgery 97, 1110–1118.

Paterson, H., Mander, B., Muir, P., Phillips, H., Wild, S., 2014. Deprivation and access to treatment for colorectal cancer in southeast scotland 2003–2009. Colorectal Disease 16, O51–O57.

Pearson, C., Fraser, J., Peake, M., Valori, R., Poirier, V., Coupland, V.H., Hiom, S., McPhail, S., Moffat, J., Lyratzopoulos, G., et al., 2019. Establishing population-based surveillance of diagnostic timeliness using linked cancer registry and administrative data for patients with colorectal and lung cancer. Cancer epidemiology 61, 111–118.

Public Health Scotland, 2020. Cancer incidence in scotland (to december 2018. https://beta.isdscotland.org/find-publications-and-data/conditions-and-diseases/cancer/cancer-incidence-in-scotland/28-april-2020/.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.C., Saunders, L.D., Beck, C.A., Feasby, T.E., Ghali, W.A., 2005. Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. Medical care , 1130–1139.

Robertson, R., Campbell, N.C., Smith, S., Donnan, P., Sullivan, F., Duffy, R., Ritchie, L.D., Millar, D., Cassidy, J., Munro, A., 2004. Factors influencing time from presentation to treatment of colorectal and breast cancer in urban and rural areas. British journal of cancer 90, 1479–1485.

Shack, L., Rachet, B., Brewster, D., Coleman, M., 2007. Socioeconomic inequalities in cancer survival in scotland 1986–2000. British journal of cancer 97, 999–1004.

Smith, J., Tilney, H., Heriot, A., Darzi, A., Forbes, H., Thompson, M., Stamatakis, J., Tekkis, P., 2006. Social deprivation and outcomes in colorectal cancer. British journal of surgery 93, 1123–1131.

Taylor, J.C., Swinson, D., Seligmann, J.F., Birch, R.J., Dewdney, A., Brown, V., Dent, J., Rossington, H.L., Quirke, P., Morris, E.J., et al., 2021. Addressing the variation in adjuvant chemotherapy treatment for colorectal cancer: Can a regional intervention promote national change? International Journal of Cancer 148, 845–856.

Yang, Y., Wang, G., He, J., Ren, S., Wu, F., Zhang, J., Wang, F., 2017. Gender differences in colorectal cancer survival: A meta-analysis. International journal of cancer 141, 1942–1949.

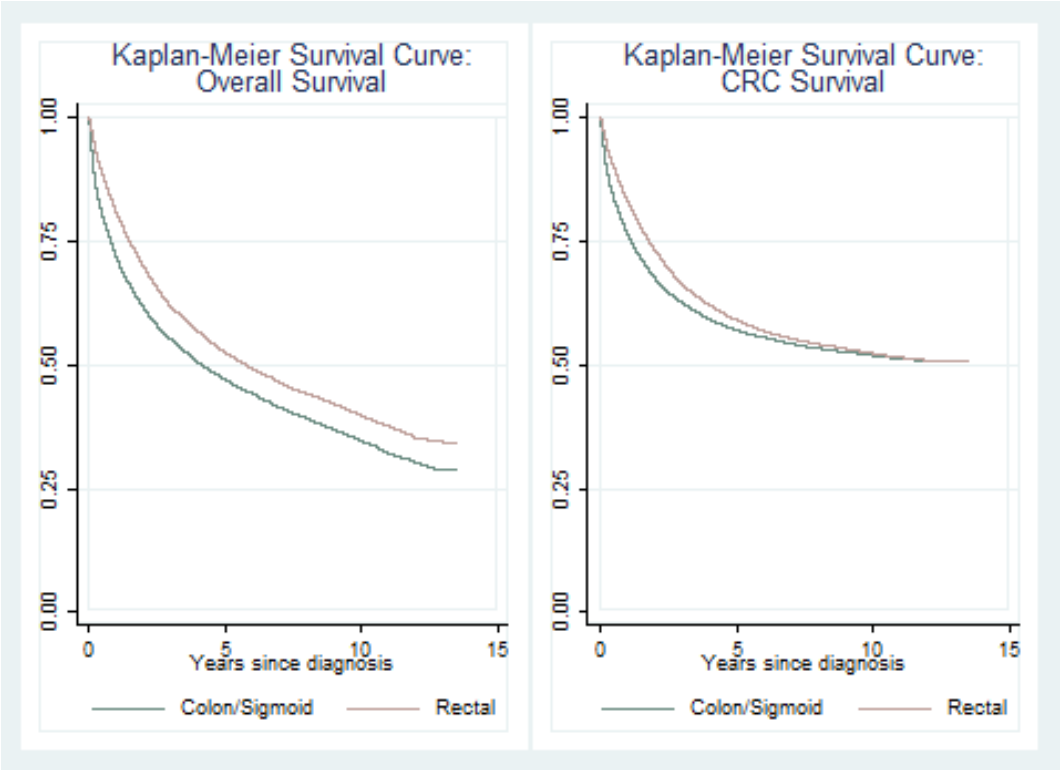

Figure 2: Kaplan-Meier Survival Curves by CRC types

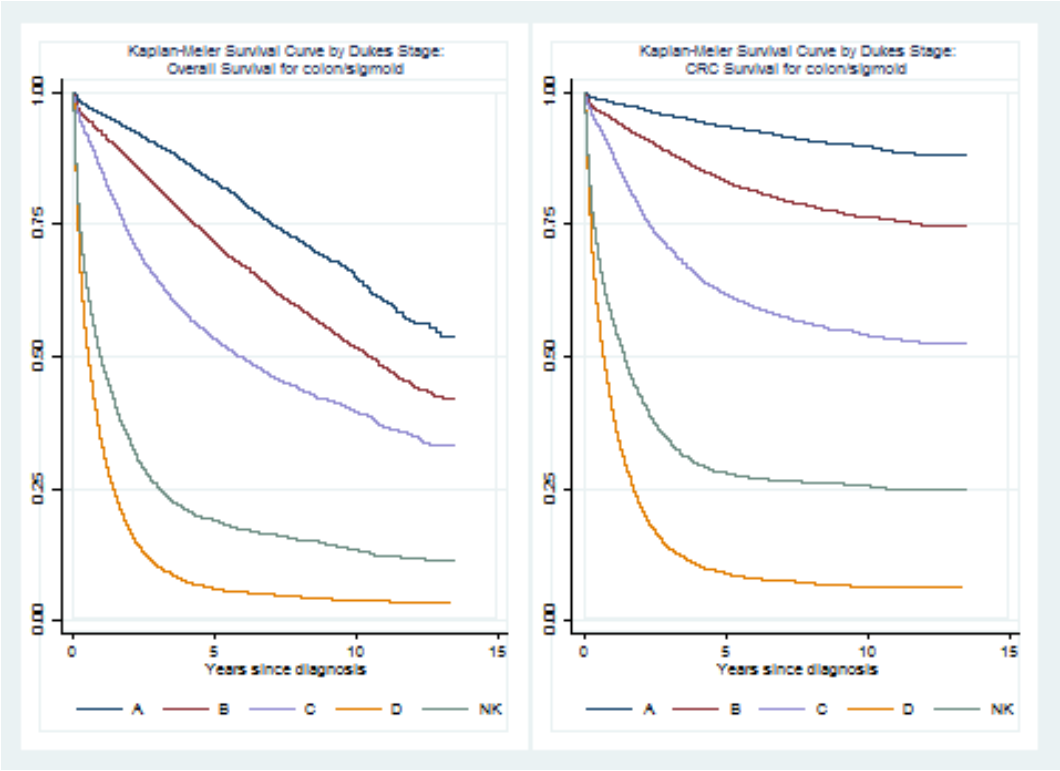

Figure 3: Kaplan-Meier Survival Curves for Colon patients by Dukes's Stage

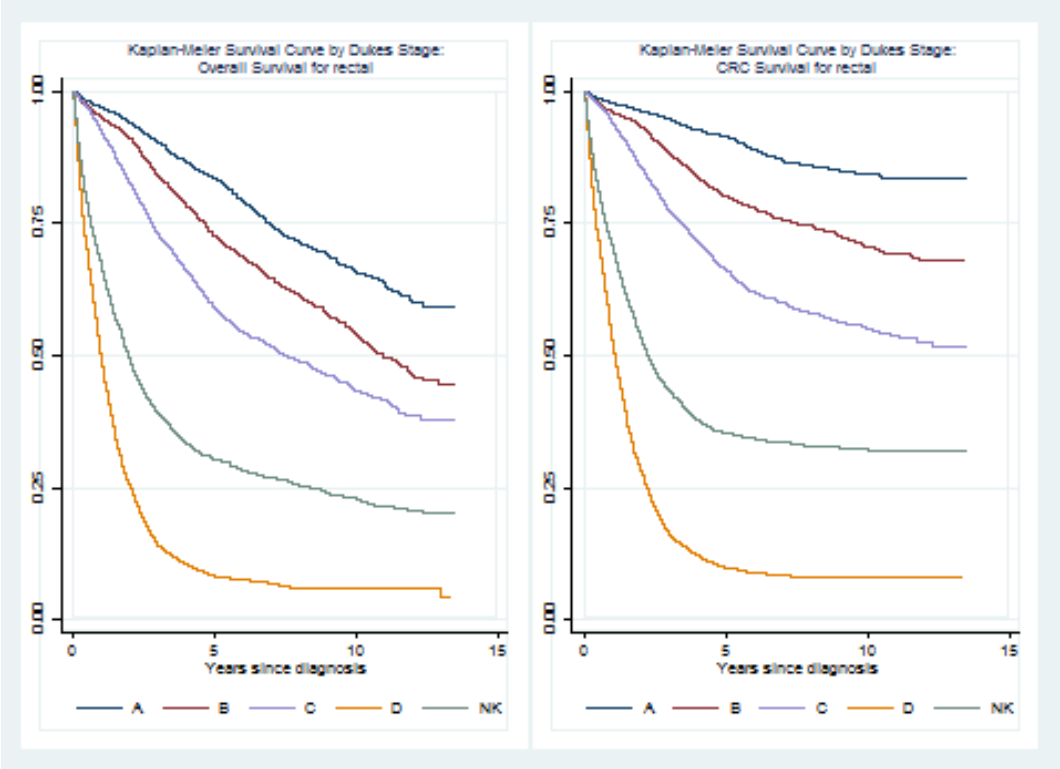

Figure 4: Kaplan-Meier Survival Curves for Rectal patients by Dukes's Stage